下载APP

12｜空间音频入门：如何实现“声临其境”？

2021-12-17 冯建元

《搞定音频技术》 课程介绍 >

讲述：冯建元

时长 12:09 大小 11.15M

你好，我是建元。

前面的课程，我们主要介绍了音频链路中必备的前处理模块，可以说是音频体验的基础设施。在这个基础上我们就可以实现很多有趣的应用，也就是音频应用的上层建筑。这节课，我们先来讲一讲如何重现一个声场环境，让音频能够穿越时空的限制，做到“声临其境”的。

你可以先闭上眼睛，尝试听一听周围的声音。你是不是可以感受到有的声音是从左边传来，有的声音会从右边，甚至是从后面传来。我们如果身处地铁、小巷、教室或者球场，是不是听到的声音也会有所不同。有的时候，浴室里的声音会比较浑浊，而操场等空旷的场景声音又会显得比较干燥？其实这些都是我们人耳对空间的感知。

所谓空间音频的技术，就是把现实中这些对声音的感知，能够用空间音频采集设备和播放设备还原出来。空间音频涉及空间声学、空间声采集、空间声重放等细分领域，内容比较多。这里我会按照空间音频的基本原理以及空间音频的采集和播放，结合已有的解决方案，来给你讲一讲空间音频背后的原理和使用方法，从而让你能够快速步入空间音频这一音频“元宇宙”的入口。

方位判断与双耳效应

好的，刚才我们说过，我们可以感知到声音是从不同的方向传来的。我们不妨先来看看这所谓的“方向感”是怎么产生的。

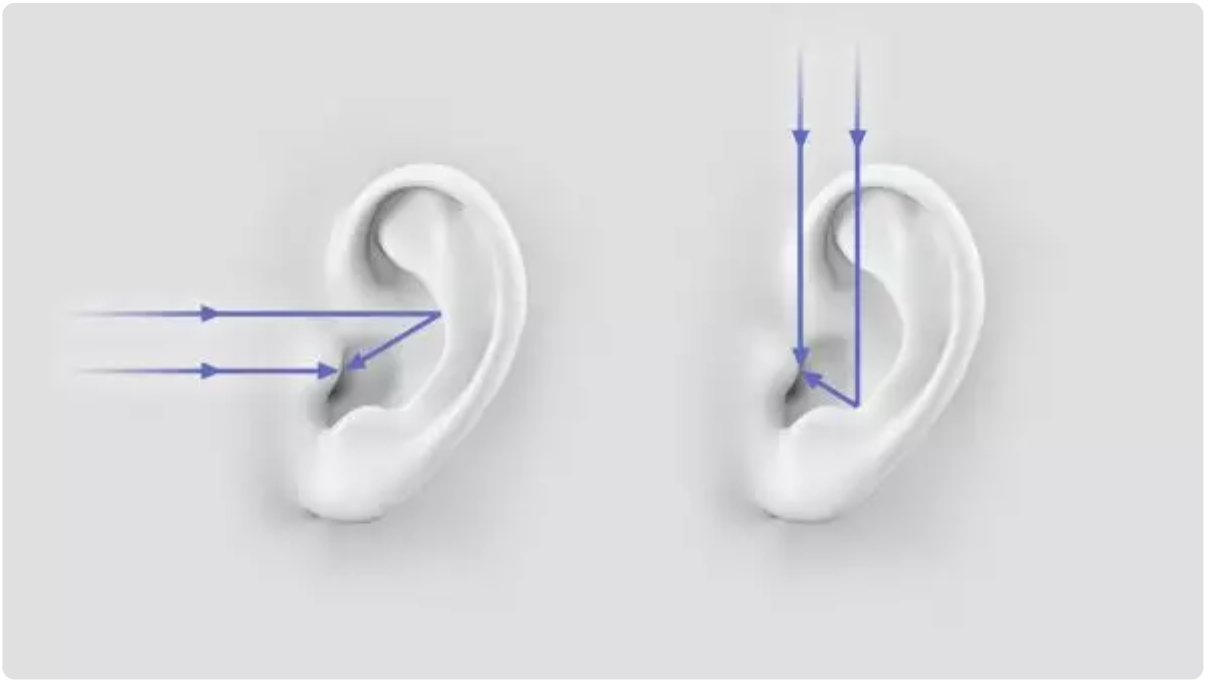

图1 耳廓收音示意图

我们可以通过图 1 看到人耳的耳廓在接收不同方向的音源时，会让声波以不同的路径传导至内耳。这样，不同方向的声波传输到内耳的时候，音色就会由于耳廓的形状而产生各向异性。除此之外，由于我们有两个耳朵，所以音源在不同方向时声波到达耳朵的时间也会不同，这一点我们可以结合图 2 来理解一下。

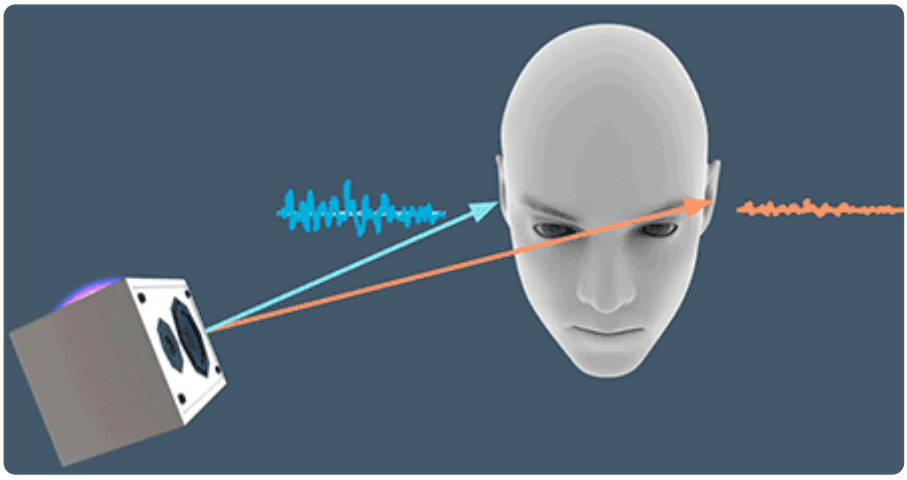

图2 双耳效应示意图

其实很简单，如果音源在你的左侧，那么左耳会先接收到声波；相反如果音源在右侧，右耳会先收到声音。同时由于人的头部也会对声音的传播产生影响，如果音源在左侧，那么声波需要越过头部这个“障碍”才能传递到右耳，那么相对于左耳，音色和能量可能都会有所衰减。**这其实就是空间音频里常说的“双耳效应”，即依靠双耳间的音量差、时间差和音色差来判别声音方位的效应。**

距离感和空间感

好的，知道了方向感是怎么产生的，我们再看看我们是如何听出音源所在的不同距离的。距离感，给人的第一感觉是：如果这个声音的音量小，那么一定是因为它离我们比较远，而声音大则是距离近。

而实际上声音的大小是相对的，比如可以音源离你很近但却只是低声细语，或者离得很远但用一个功率比较大的音箱，大声播放。所以用音量本身来判断声音的距离是不够的。**实际上我们人耳对距离的感知是相对的。**比如声音播放时音量由小变大，我们会感觉声音在靠近。或者同时播放两个吉他的声音，你会感觉声音小的是在远一些的位置。

除了音量之外，还由于声波在空气传播中会产生衰减。而其中高频的声音衰减快、低频的衰减慢。那么同一个音量的声音，如果高频比较多，我们会觉得它离你更近一些。如图 3 所示，低频的声音可以传播得更远，而 8kHz 以上的声音如果超过 1 千米就很难被听到了。

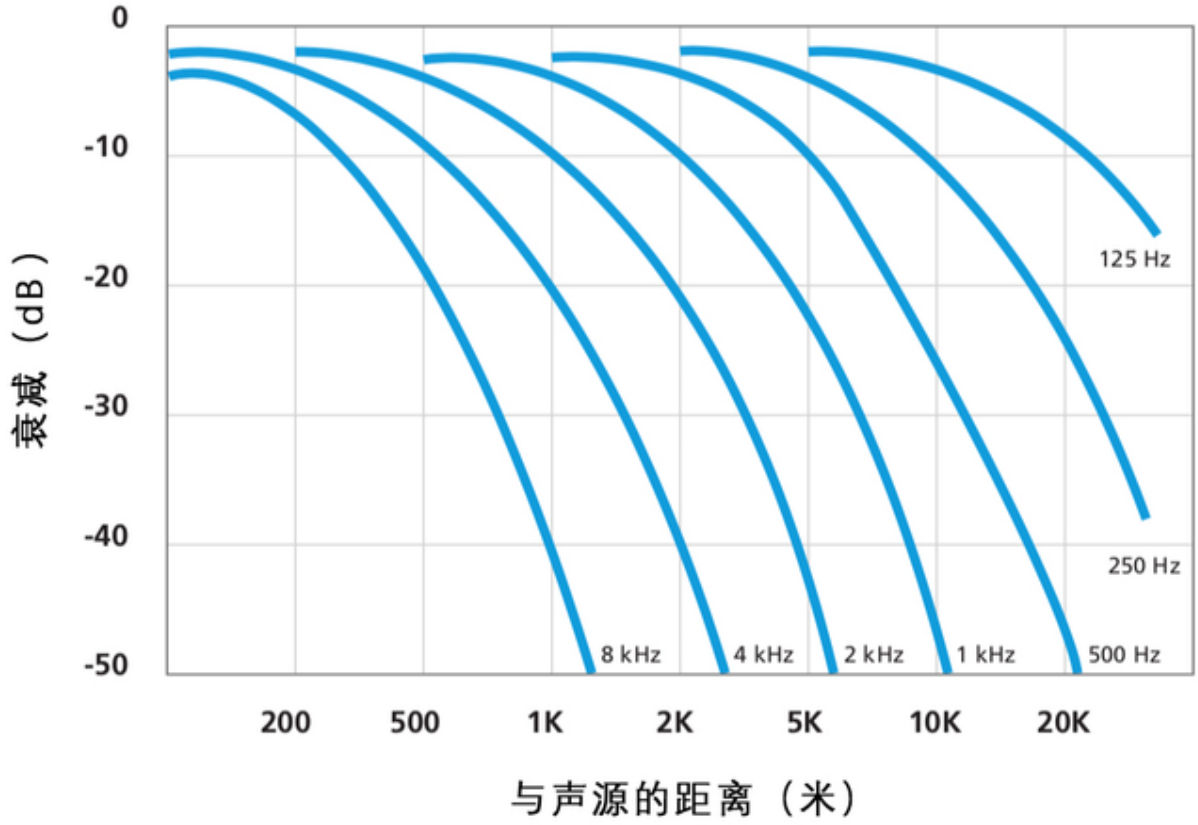

图3 不同频率的声波空气衰减曲线

除了方向和距离，在之前讲回声消除的时候我们提到过混响的概念。**其实混响也是我们感知声场空间大小的重要一环。**简单地说，大房间的混响持续时间长，而小房间的混响持续时间短。我们能通过声音的混响来感知所在空间的大小。当然混响还和房间墙壁的材料、形状以及房间中障碍物的情况有关，所以我们对空间的感知相比方向和距离没有那么准确。

空间音频的采集

好了，有了空间音频的基础知识，让我们来看看我们是如何把空间音频采集下来的。

入耳式麦克风与人工头

很显然想要把人耳听到的声音完整的保留，我们可以采用入耳式麦克风，直接把左右耳道接收到的音频给录下来。或者如图 4 所示，可以使用人工头的方式，通过仿生模型，构建人头和耳廓、耳道等部位，然后**通过人工头上的人工耳中内置的麦克风来采集空间音频。**

图4 人工头示意图

入耳式麦克风和人工头采集的区别其实也显而易见的。如果你用入耳式麦克风采集的音频再用入耳式耳机播放，那么基本上可以做到完美还原。而如果是用人工头录制，那么由于耳廓的形状、头的形状等都和你自己有所不同，所以虽然可以做到很大程度上的空间还原，但和你自己实际到场景中去听，还是有一些差别的。

在实际使用中，每个人的耳朵、头的形状都不一样，但大体的形状和位置是相同的。所以**利用人工头做音频录制在很多影视和游戏音频制作中会经常用到。**

Ambisonics

但拿人工头或者入耳式麦克风采集到的音频都还只是固定方向的立体声还原，且只能还原采集时人头朝向的声音。如果想把整个空间的声场都录下来，从而在回放的时候，你可以转动自己的头聆听任意方向的声音，那么就需要另一套叫做高保真度立体声像复制（Ambisonics）的技术。

高保真度立体声像复制源于 20 世纪 70 年代牛津大学的一个三维空间声场重构技术研究。技术的核心是将远端中能听到的声音通过特制的麦克风录制，比如一阶 Ambisonics 麦克

风（由四个完全相同的麦克风单元构成一个立方体阵列）的方式复制下来，我在图 5 里展示了几个常见的一阶 Ambisonic 麦克风。

图5 一阶Ambisonic麦克风

这里由一阶 Ambisonic 麦克风采集的原始数据我们叫 A-format，是无法直接播放的，需要按照多通道转码格式先转为 4 通道的 B-format。4 通道的 B-format 也叫作一阶 B-format。其中的四个通道分别称为 W、X、Y 和 Z。简单一点来理解，这四个方向分别代表了一个球形声场的中心、左右、前后和上下。B-format 的数据就可以用软件渲染成任意播放设备支持的格式，比如立体声、2.1、5.1 甚至 7.1。

低阶的 Ambisonics 麦克风可以还原一个比较小的声场，而如果是飞机场、大型演唱会等场景，则可能需要一个如图 6 所示的高阶的 Ambisonic 麦克风。我们可以看到阶数越多需要的麦克风的个数也就越多。

图6 高阶Ambisonic麦克风

Ambisonics 技术在 AR、VR 等需要转动视角的场景里可以很好地还原整个声场的听感，所以被广泛应用。

空间音频的播放

好的，讲完了空间音频的采集。我们再来看看我们是如何把采集到的声音回放出来的。空间音频最常用的方式就是使用耳机播放。在空间音频的原理部分我们说过，要想感受空间音频最少需要一个双声道的音频播放，来让左右耳感知音量、音色、时间延迟等差异，从而形成空间感。同时由于耳机的播放单元离人耳比较近，无需引入额外的声波传递带来听感的变化。所以**用耳机来还原空间音频相对比较准确。**

但是耳机由于受限于结构和功率的限制，在低音部分的表现可能就不如音箱来得“震撼”。而且如果需要同时给多人播放体验空间音频，多声道的音箱系统会是我们的另一条可选的播放途径。多通道音箱系统的渲染格式经过多年的发展已经比较标准化了。常见的多通道音箱系统有 2.0、2.1、5.1、5.1.2、7.1、7.1.2 等等。那么它们分别代表什么含义呢？

其实这里我们可以把数字分成 A.B.C 三个部分。其中，A 代表有多少个环绕声扬声器（前置、中置和环绕）的数量；B 表示有多少个超低音音箱；C 代表顶部或向上发声扬声器的数量。

举个例子，如图 7 所示，这是一个 Dolby 7.1.4 声道的全景声音箱系统。其中环绕扬声器是其中长方体表示的左前、中置、右前、左、右、左后、右后共 7 个环绕扬声器。超低音音箱是左上方的正方形，它一般都是用于增强 200Hz 以下的超低频率的声音，并且由于低频部分的声波波长较长，在房间内方向的感知不明显，所以方位上来说只需要放在房间中的任意位置即可。而在顶部的扬声器一般在听音位置也就是图 7 沙发的上方均匀排布（黑色圆形），用于提供来自上方的声音渲染。

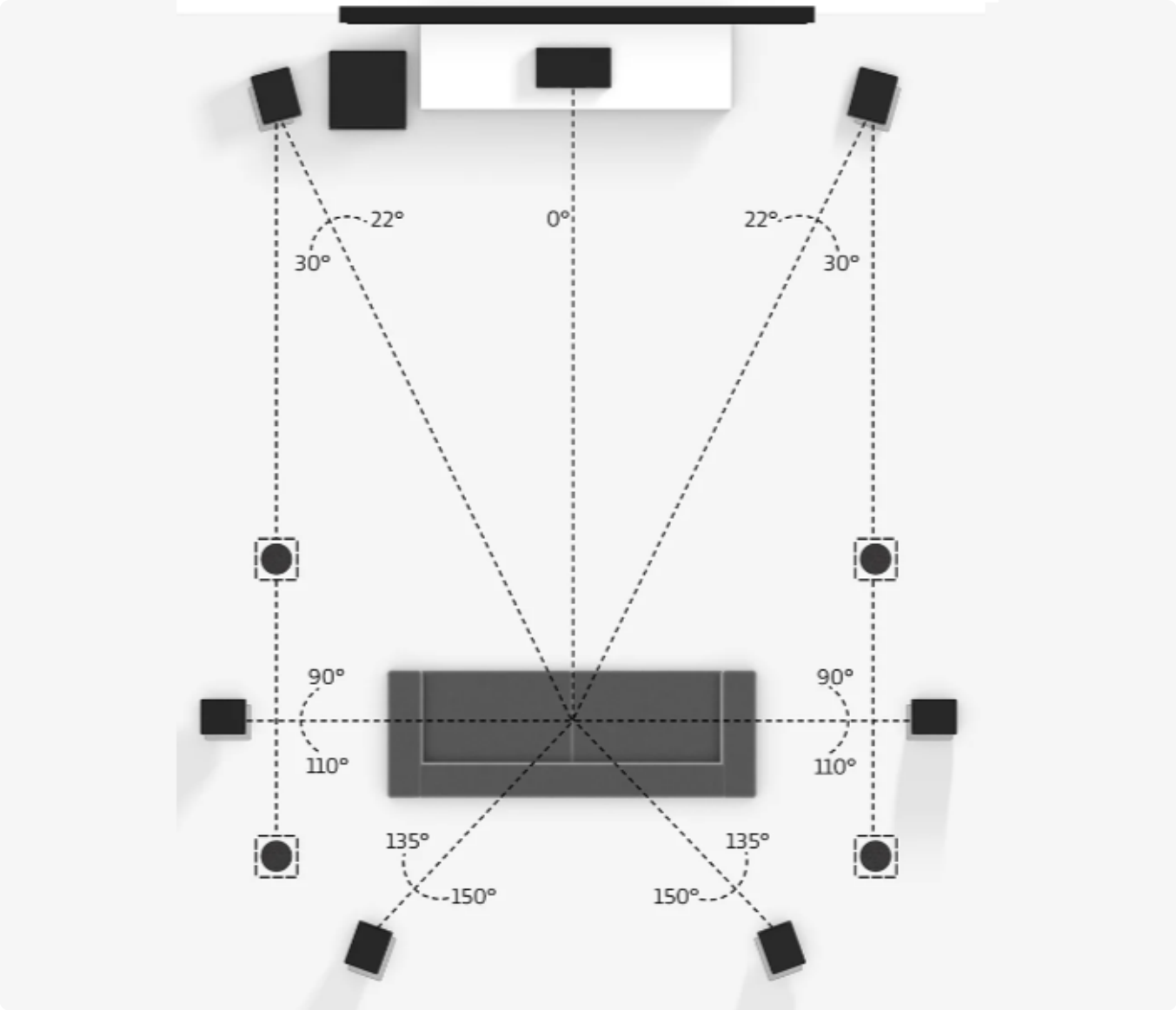

图7 Dolby7.1.4家庭影院系统

我们可以看到，**不同的声音借由各个扬声器的位置和播放渲染可以实现比较好的声场还原。**但是这种传统的家庭影院的布局并不能完美地还原声场。比如来自下方的声音就不能

清晰重现，并且听音者只能在相对固定的位置（比如沙发）上才能获得正确的听感。所以这样的家庭影院看看电影还是不错的，但是玩一些 VR 游戏之类的就会显得声场渲染不足。

其实在普通的电影院也只是这种家庭影院的扩展，只是多一些环绕音箱让声音分辨力更强等等。从原理上来说音箱分布得越多、越密集，对声场的还原就越好。但显然这样所需的成本是巨大的，这其实也是多声道影院系统推出这么多年，推广一直受限的主要原因。

小结

好的，今天的课到这里就要结束了。让我们来总结一下这节课的内容。

首先，我们对空间内声音的位置的感知是因为我们的耳廓形状和双耳效应。对空间中不同位置传来的声音，左右耳会得到不同的音色、音量以及延迟变化。距离感则主要是声音的相对音量大小以及声波空气衰减时，高频衰减快于低频导致的频率能量变化来感知。而听混响的持续时间则是我们感知所处空间大小的方法。

空间音频的采集可以使用入耳式麦克风或者人工头的形式来录制可用耳机回放的立体声。而 Ambisonics 技术则可以采用 Ambisonic 麦克风来对整体空间声场进行采集，然后通过后期的渲染来自由调整听音的方向。

这里我想补充一下，在影视制作或者现场录制的时候你可能会采用一些所谓的立体声录制麦克风。比如常见的 A-B 式麦克风、X-Y 式麦克风等，这些录制方式可以采集双声道的立体声信号。这些声音是双声道，虽然你可以听出一些方位感和空间感，但是也可以作为简易的现场录音的方法的，只是这些立体声的录制无法通过后续播放来精准还原现场的声场。

所以在这节课里我没有把它归入空间音频的录制方法里。但这些立体声麦克风采集的音频，回放空间听感效果还是会优于单声道麦克风，所以在无需完美还原的场景里，也是可以使用的。

最后，想要感受空间音频逼真的方向感和空间感我们可以使用耳机、环绕声或者全景声音箱的形式来进行回放。其实单声道的声音经过空间音频算法的渲染也可以让人有“声临其境”的感觉，我会在下节课为你详细讲解。

思考题

好的，这里给你留一个思考题。

我们在使用蓝牙耳机来听空间音频的时候，为什么有的时候会失去空间感呢？这里提示一下蓝牙耳机有很多协议，其中有的是单向高保真音频协议，例如 A2DP、LDAC 等；有的是双向低保真音频协议，例如 HFP（Hands-Free Profile）或 HSP（HeadSet Profile）。

你可以把你的答案和疑惑写下来，分享到留言区，与我一起讨论。我们下节课再见。

分享给需要的人，Ta订阅后你可得 20 元现金奖励

生成海报并分享

赞 0 提建议

© 版权归极客邦科技所有，未经许可不得传播售卖。 页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 11 | 网络差怎么办？音频网络传输与抗弱网策略

精选留言

写留言

> 由作者筛选后的优质留言将会公开显示，欢迎踊跃留言。